

#15 PCT09

RAW SEQUENCE LISTING
PATENT APPLICATION: US/09/869,142

DATE: 07/09/2001
TIME: 10:10:43

Input Set : A:\PTO.txt
Output Set: N:\CRF3\07092001\I869142.raw

3 <110> APPLICANT: SHOWA DENKO K.K.
5 <120> TITLE OF INVENTION: NOVEL RHODOCOCCLUS BACTERIA, NITRILASE GENE, NITRIL HYDRATASE
GENE AND

6 AMIDASE GENE FROM RHODOCOCCLUS BACTERIUM, AND PROCESS FOR PRODUCING CARBOXYLIC
7 ACIDS USING THEM
9 <130> FILE REFERENCE: Q64574

C--> 11 <140> CURRENT APPLICATION NUMBER: US/09/869,142

C--> 11 <141> CURRENT FILING DATE: 2001-06-26

11 <150> PRIOR APPLICATION NUMBER: USSN 60/183,754

12 <151> PRIOR FILING DATE: 2000-02-22

14 <150> PRIOR APPLICATION NUMBER: USSN 60/183,821

15 <151> PRIOR FILING DATE: 2000-02-22

17 <150> PRIOR APPLICATION NUMBER: JPA 2000-107855

18 <151> PRIOR FILING DATE: 2000-04-10

20 <150> PRIOR APPLICATION NUMBER: JPA 2000-021797

21 <151> PRIOR FILING DATE: 2000-01-26

23 <150> PRIOR APPLICATION NUMBER: JPA 11-303212

24 <151> PRIOR FILING DATE: 1999-10-26

26 <160> NUMBER OF SEQ ID NOS: 7

28 <170> SOFTWARE: PatentIn version 3.1

30 <210> SEQ ID NO: 1

31 <211> LENGTH: 1531

32 <212> TYPE: DNA

33 <213> ORGANISM: Rhodococcus sp.

35 <220> FEATURE:

36 <221> NAME/KEY: exon

37 <222> LOCATION: (324)..(1421)

38 <223> OTHER INFORMATION:

41 <400> SEQUENCE: 1

42 agcttgacca tgattacgaa ttcgagctcg gtacccgggg atcgaaccag caacggggac 60

44 gcacagtcga cgtagacctc gacctatccg ccgttcgcga gaaggacacc gaccaccacc 120

46 acttcaacat ccttcaacgt gcccgccag tccttcgacg aatcgaaacg gcgaagagcc 180

48 gcctcggacc ccccgccga accgctcgat gaactcccct acacgggtgg cgcagaatgc 240

50 caggaccctg gtcattccac gtcaattcac gcgccttttc acctcgact gtctgcgcaa 300

52 acacaagcaa cggaggtacg gac atg gtc gaa tac aca aac aca ttc aaa gtt 353

53 Met Val Glu Tyr Thr Asn Thr Phe Lys Val 10

54 1 5 401

56 gct gcg gtg cag gca cag cct gtg tgg ttc gac gcg gcc aaa acg gtc 401

57 Ala Ala Val Gln Ala Gln Pro Val Trp Phe Asp Ala Ala Lys Thr Val 25

58 15 20 25 449

60 gac aag acc gtg tcc atc atc gcg gaa gca gcc cgg aac ggg tgc gag 449

61 Asp Lys Thr Val Ser Ile Ile Ala Glu Ala Ala Arg Asn Gly Cys Glu 40

62 30 35 40 497

64 ctc gtt gcg ttt ccc gag gta ttc atc ccg ggg tac ccg tac cac atc 497

65 Leu Val Ala Phe Pro Glu Val Phe Ile Pro Gly Tyr Pro Tyr His Ile 55

66 45 50 55 545

68 tgg gtc gac agc ccg ctc gcc gga atg gcg aag ttc gcc gtg cgc tac 545

69 Trp Val Asp Ser Pro Leu Ala Gly Met Ala Lys Phe Ala Val Arg Tyr

ENTERED

RAW SEQUENCE LISTING

PATENT APPLICATION: US/09/869,142

DATE: 07/09/2001

TIME: 10:10:43

Input Set : A:\PTO.txt

Output Set: N:\CRF3\07092001\I869142.raw

70	60	65	70	593
72	cac gag aat tcc ctg acg atg gac agc ccg cac gta cag cgg ttg ctc			
73	His Glu Asn Ser Leu Thr Met Asp Ser Pro His Val Gln Arg Leu Leu			
74	75	80	85	90
76	gat gcc gcc cgc gac cac aac atc gcc gta gtg gtg gga atc agc gag			641
77	Asp Ala Ala Arg Asp His Asn Ile Ala Val Val Val Gly Ile Ser Glu			
78	95	100	105	
80	cgg gat ggc ggc agc ttg tac atg acc cag ctc atc atc gac gcc gat			689
81	Arg Asp Gly Gly Ser Leu Tyr Met Thr Gln Leu Ile Ile Asp Ala Asp			
82	110	115	120	
84	ggg caa ctg gtc gcc cga cgc cgc aag ctc aag ccc acc cac gtc gag			737
85	Gly Gln Leu Val Ala Arg Arg Arg Lys Leu Lys Pro Thr His Val Glu			
86	125	130	135	
88	cgt tcg gta tac gga gaa gga aac ggc tcg gat atc tcc gtg tac gac			785
89	Arg Ser Val Tyr Gly Glu Gly Asn Gly Ser Asp Ile Ser Val Tyr Asp			
90	140	145	150	
92	atg cct ttc gca cgg ctt ggc gcg ctc aac tgc tgg gag cat ttc cag			833
93	Met Pro Phe Ala Arg Leu Gly Ala Leu Asn Cys Trp Glu His Phe Gln			
94	155	160	165	170
96	acg ctc acc aag tac gca atg tac tcg atg cac gag cag gtg cac gtc			881
97	Thr Leu Thr Lys Tyr Ala Met Tyr Ser Met His Glu Gln Val His Val			
98	175	180	185	
100	gcg agc tgg cct ggc atg tcg ctg tac cag ccg gag gtc ccc gca ttc			929
101	Ala Ser Trp Pro Gly Met Ser Leu Tyr Gln Pro Glu Val Pro Ala Phe			
102	190	195	200	
104	ggt gtc gat gcc cag ctc acg gcc acg cgt atg tac gca ctc gag gga			977
105	Gly Val Asp Ala Gln Leu Thr Ala Thr Arg Met Tyr Ala Leu Glu Gly			
106	205	210	215	
108	caa acc ttc gtg gtc tgc acc acc cag gtg gtc aca ccg gag gcc cac			1025
109	Gln Thr Phe Val Val Cys Thr Thr Gln Val Val Thr Pro Glu Ala His			
110	220	225	230	
112	gag ttc ttc tgc gag aac gag gaa cag cga atg ttg atc ggc cga ggc			1073
113	Glu Phe Phe Cys Glu Asn Glu Glu Gln Arg Met Leu Ile Gly Arg Gly			
114	235	240	245	250
116	gga ggt ttc gcg cgc atc atc ggg ccc gac ggc cgc gat ctc gca act			1121
117	Gly Gly Phe Ala Arg Ile Ile Gly Pro Asp Gly Arg Asp Leu Ala Thr			
118	255	260	265	
120	cct ctc gcc gaa gat gag gag ggg atc ctc tac gcc gac atc gat ctg			1169
121	Pro Leu Ala Glu Asp Glu Glu Gly Ile Leu Tyr Ala Asp Ile Asp Leu			
122	270	275	280	
124	tct gcg atc acc ttg gcg aag cag gcc gct gac ccc gtg ggc cac tac			1217
125	Ser Ala Ile Thr Leu Ala Lys Gln Ala Ala Asp Pro Val Gly His Tyr			
126	285	290	295	
128	tca cgg ccg gat gtg ctg tcg ctg aac ttc aac cag cgc cgc acc acg			1265
129	Ser Arg Pro Asp Val Leu Ser Leu Asn Phe Asn Gln Arg Arg Thr Thr			
130	300	305	310	
132	ccc gtc aac acc cca ctt tcc acc atc cat gcc acg cac acg ttc gtg			1313
133	Pro Val Asn Thr Pro Leu Ser Thr Ile His Ala Thr His Thr Phe Val			
134	315	320	325	330

RAW SEQUENCE LISTING

PATENT APPLICATION: US/09/869,142

DATE: 07/09/2001

TIME: 10:10:43

Input Set : A:\PTO.txt

Output Set: N:\CRF3\07092001\I869142.raw

```

136 ccg cag ttc ggg gca ctc gac ggc gtc cgt gag ctc aac gga gcg gac 1361
137 Pro Gln Phe Gly Ala Leu Asp Gly Val Arg Glu Leu Asn Gly Ala Asp
138          335          340          345
140 gaa cag cgc gca ttg ccc tcc aca cat tcc gac gag acg gac cgg gcg 1409
141 Glu Gln Arg Ala Leu Pro Ser Thr His Ser Asp Glu Thr Asp Arg Ala
142          350          355          360
144 aca gcc acc ctc tgactcgggc gcaccctgtg gcctccgaa gcgccacggg 1461
145 Thr Ala Thr Leu
146          365
148 tgtgtgaagg ggcgagacag gggaatcgga ggatccccgg gtaccgagct cgaattcgta 1521
150 atcatggtca 1531
153 <210> SEQ ID NO: 2
154 <211> LENGTH: 366
155 <212> TYPE: PRT
156 <213> ORGANISM: Rhodococcus sp.
158 <400> SEQUENCE: 2
160 Met Val Glu Tyr Thr Asn Thr Phe Lys Val Ala Ala Val Gln Ala Gln
161 1          5          10          15
164 Pro Val Trp Phe Asp Ala Ala Lys Thr Val Asp Lys Thr Val Ser Ile
165          20          25          30
168 Ile Ala Glu Ala Ala Arg Asn Gly Cys Glu Leu Val Ala Phe Pro Glu
169          35          40          45
172 Val Phe Ile Pro Gly Tyr Pro Tyr His Ile Trp Val Asp Ser Pro Leu
173          50          55          60
176 Ala Gly Met Ala Lys Phe Ala Val Arg Tyr His Glu Asn Ser Leu Thr
177 65          70          75          80
180 Met Asp Ser Pro His Val Gln Arg Leu Leu Asp Ala Ala Arg Asp His
181          85          90          95
184 Asn Ile Ala Val Val Val Gly Ile Ser Glu Arg Asp Gly Gly Ser Leu
185          100          105          110
188 Tyr Met Thr Gln Leu Ile Ile Asp Ala Asp Gly Gln Leu Val Ala Arg
189          115          120          125
192 Arg Arg Lys Leu Lys Pro Thr His Val Glu Arg Ser Val Tyr Gly Glu
193          130          135          140
196 Gly Asn Gly Ser Asp Ile Ser Val Tyr Asp Met Pro Phe Ala Arg Leu
197 145          150          155          160
200 Gly Ala Leu Asn Cys Trp Glu His Phe Gln Thr Leu Thr Lys Tyr Ala
201          165          170          175
204 Met Tyr Ser Met His Glu Gln Val His Val Ala Ser Trp Pro Gly Met
205          180          185          190
208 Ser Leu Tyr Gln Pro Glu Val Pro Ala Phe Gly Val Asp Ala Gln Leu
209          195          200          205
212 Thr Ala Thr Arg Met Tyr Ala Leu Glu Gly Gln Thr Phe Val Val Cys
213          210          215          220
216 Thr Thr Gln Val Val Thr Pro Glu Ala His Glu Phe Phe Cys Glu Asn
217 225          230          235          240
220 Glu Glu Gln Arg Met Leu Ile Gly Arg Gly Gly Gly Phe Ala Arg Ile
221          245          250          255
224 Ile Gly Pro Asp Gly Arg Asp Leu Ala Thr Pro Leu Ala Glu Asp Glu

```

RAW SEQUENCE LISTING
PATENT APPLICATION: US/09/869,142

DATE: 07/09/2001
TIME: 10:10:43

Input Set : A:\PTO.txt
Output Set: N:\CRF3\07092001\I869142.raw

225		260		265		270	
228	Glu Gly Ile	Leu Tyr Ala Asp	Ile Asp Leu Ser Ala	Ile Thr Leu Ala			
229		275		280		285	
232	Lys Gln Ala	Ala Asp Pro Val	Gly His Tyr Ser Arg	Pro Asp Val Leu			
233		290		295		300	
236	Ser Leu Asn Phe	Asn Gln Arg Arg	Thr Thr Pro Val	Asn Thr Pro Leu			
237	305		310		315		
240	Ser Thr Ile His	Ala Thr His Thr	Phe Val Pro Gln	Phe Gly Ala Leu			
241		325		330		335	
244	Asp Gly Val Arg	Glu Leu Asn Gly	Ala Asp Glu Gln	Arg Ala Leu Pro			
245		340		345		350	
248	Ser Thr His Ser	Asp Glu Thr Asp	Arg Ala Thr Ala	Thr Leu			
249		355		360		365	
252	<210> SEQ ID NO: 3						
253	<211> LENGTH: 2822						
254	<212> TYPE: DNA						
255	<213> ORGANISM: Rhodococcus sp.						
257	<220> FEATURE:						
258	<221> NAME/KEY: CDS						
259	<222> LOCATION: (1379)..(2068)						
260	<223> OTHER INFORMATION: nitrile hydratase beta subunit						
263	<220> FEATURE:						
264	<221> NAME/KEY: CDS						
265	<222> LOCATION: (2082)..(2693)						
266	<223> OTHER INFORMATION: nitrile hydratase alpha subunit						
269	<400> SEQUENCE: 3						
270	ctagaggatc	tcggtcatcg	cgataccatc	gttgccggacg	atgatgtcca	atacgtacca	60
272	ctggtccgcg	gtcaacttct	cttgatcgac	cacgttatgg	attctacgac	tcagggaccg	120
274	gctcacggct	tccaggggcg	ctccgaccaa	aggatgatcg	acgacatttc	cggattcagc	180
276	caccgcttcc	gactcgatca	ttcctgtccc	tccccgtcca	cgcgcatgtg	atcttacctc	240
278	ctcatcaaga	ggatatccac	tgaacgaatt	atttcaagtg	gaagtacttg	gagtcgatcc	300
280	tacacgtgag	tggacgatgc	ctggggcgcta	gtcggatgtg	caaccacccc	acccccctct	360
282	cccgccctacg	ccgaagaccg	gaaccggcgt	cgtccctgcc	tgccgtctct	ggcaactggt	420
284	gtgaacgccc	gagcggccct	cacggctctt	cagttggcgc	ggatcgccat	ggcggacgtc	480
286	gcccacggcg	ggacctacgc	atcttcggcc	ggaaggcagc	cgcggtcacg	aacacctagc	540
288	ggcagtcgag	cacctgagac	gaaggccgcc	ggcgctcctg	cccgaaaatc	cgcagcccag	600
290	ccgtgacagc	caacagtcgt	ggcggttccc	tccccctcta	gggtctttga	ctcggcgcca	660
292	acgcctgcga	gggcgctcgt	cgcgaccac	ttgtcgaggt	cgggtgccga	cgtcaccgag	720
294	cgcacccttc	ttcgtgctct	gcgcacggcg	ccggaccgcg	accgcggcaa	cactacgacg	780
296	tctgacaatg	ctgatccctt	gccgcccggc	ttggacgacc	acagttgcta	cgagcatgcg	840
298	gagccaacca	taggcatcat	gcgatcgccg	gagtcttcat	cctatttttg	gatgcgcagg	900
300	attaacacat	ctacacattg	acatccgttc	cgatgtgaag	taaaaattgt	cacgtagggc	960
302	ggcaggcgaa	gtctgcagct	cgaacatcga	agggtgggag	ccgagagatc	ggagacgcag	1020
304	acaccgggag	ggaacttagc	ctcccgacc	gatgcgtgtc	ctggcaacgc	ctcaagattc	1080
306	agcgcaagcg	attcaatctt	gttacttcca	gaaccgaatc	acgtccccgt	agtgtgcggg	1140
308	gagagcgccc	gaacgcaggg	atggttatcca	tgcgccccct	ctcttttcga	acgagaaccg	1200
310	gccggtacag	tcaatccgga	cacattgtga	cgccgttcaa	cgattgttgt	gctgtgaagg	1260
312	attcactcaa	gccaaactgat	atcgccattc	cgttgccgga	acatttgacg	ccttctccct	1320
314	acgagtagaa	gccagctgga	ccctctttga	gcccgactcc	gatgaaagga	atgaggaa	1378

RAW SEQUENCE LISTING
PATENT APPLICATION: US/09/869,142

DATE: 07/09/2001
TIME: 10:10:43

Input Set : A:\PTO.txt
Output Set: N:\CRF3\07092001\I869142.raw

316	atg	gat	ggt	atc	cac	gac	aca	ggc	ggc	atg	acc	gga	tac	gga	ccg	gtc	1426
317	Met	Asp	Gly	Ile	His	Asp	Thr	Gly	Gly	Met	Thr	Gly	Tyr	Gly	Pro	Val	
318	1			5					10						15		
320	ccc	tat	cag	aag	gac	gag	ccc	ttc	ttc	cac	tac	gag	tgg	gag	ggt	cga	1474
321	Pro	Tyr	Gln	Lys	Asp	Glu	Pro	Phe	Phe	His	Tyr	Glu	Trp	Glu	Gly	Arg	
322				20				25						30			
324	acc	ctg	tcg	att	ctg	acc	tgg	atg	cat	ctc	aag	ggc	atg	tcg	tgg	tgg	1522
325	Thr	Leu	Ser	Ile	Leu	Thr	Trp	Met	His	Leu	Lys	Gly	Met	Ser	Trp	Trp	
326				35				40					45				
328	gac	aag	tcg	cgg	ttc	ttc	cgg	gag	tcg	atg	ggg	aac	gaa	aac	tac	gtc	1570
329	Asp	Lys	Ser	Arg	Phe	Phe	Arg	Glu	Ser	Met	Gly	Asn	Glu	Asn	Tyr	Val	
330		50					55				60						
332	aac	gag	att	cgc	aac	tcg	tac	tac	acc	cac	tgg	ctg	agt	gcg	gcg	gaa	1618
333	Asn	Glu	Ile	Arg	Asn	Ser	Tyr	Tyr	Thr	His	Trp	Leu	Ser	Ala	Ala	Glu	
334	65					70					75				80		
336	cgt	atc	ctc	gtc	gcc	gac	aag	atc	atc	acc	gaa	gaa	gag	cga	aag	cac	1666
337	Arg	Ile	Leu	Val	Ala	Asp	Lys	Ile	Ile	Thr	Glu	Glu	Glu	Arg	Lys	His	
338				85					90					95			
340	cgc	gtg	cag	gag	atc	ctc	gag	ggt	cgg	tac	acg	gac	agg	aac	ccg	tcg	1714
341	Arg	Val	Gln	Glu	Ile	Leu	Glu	Gly	Arg	Tyr	Thr	Asp	Arg	Asn	Pro	Ser	
342				100				105						110			
344	cgg	aag	ttc	gat	ccg	gcc	gag	atc	gag	aag	gcg	atc	gag	agg	ctt	cac	1762
345	Arg	Lys	Phe	Asp	Pro	Ala	Glu	Ile	Glu	Lys	Ala	Ile	Glu	Arg	Leu	His	
346			115				120					125					
348	gag	ccc	cac	tcc	cta	gtg	ctt	cca	gga	gcg	gag	ccg	agt	ttc	tcc	ctc	1810
349	Glu	Pro	His	Ser	Leu	Val	Leu	Pro	Gly	Ala	Glu	Pro	Ser	Phe	Ser	Leu	
350		130					135				140						
352	ggt	gac	aag	gtc	aaa	gtg	aag	aac	atg	aac	ccg	ctg	gga	cac	aca	cgg	1858
353	Gly	Asp	Lys	Val	Lys	Val	Lys	Asn	Met	Asn	Pro	Leu	Gly	His	Thr	Arg	
354	145				150						155				160		
356	tgc	ccg	aag	tat	gtg	cgg	aac	aga	atc	ggg	gaa	atc	gtc	acc	tcc	cac	1906
357	Cys	Pro	Lys	Tyr	Val	Arg	Asn	Arg	Ile	Gly	Glu	Ile	Val	Thr	Ser	His	
358				165					170					175			
360	ggg	tgc	cag	atc	tat	ccc	gag	agc	agc	tcc	gcc	ggc	ctc	ggc	gac	gat	1954
361	Gly	Cys	Gln	Ile	Tyr	Pro	Glu	Ser	Ser	Ser	Ala	Gly	Leu	Gly	Asp	Asp	
362			180					185					190				
364	ccc	cgc	ccg	ctc	tac	acg	gtc	gcg	ttt	tcc	gcc	cag	gaa	ctg	tgg	ggc	2002
365	Pro	Arg	Pro	Leu	Tyr	Thr	Val	Ala	Phe	Ser	Ala	Gln	Glu	Leu	Trp	Gly	
366			195				200					205					
368	gac	gac	gga	aac	ggg	aaa	gac	gta	gtg	tgc	gtc	gat	ctc	tgg	gaa	ccg	2050
369	Asp	Asp	Gly	Asn	Gly	Lys	Asp	Val	Val	Cys	Val	Asp	Leu	Trp	Glu	Pro	
370		210				215				220							
372	tac	ctg	atc	tct	gcg	tga	aaggaatac	g	ata	gtg	agc	gag	cac	gtc	aat		2099
373	Tyr	Leu	Ile	Ser	Ala					Val	Ser	Glu	His	Val	Asn		
374	225								230					235			
376	aag	tac	acg	gag	tac	gag	gca	cgt	acc	aag	gca	atc	gaa	acc	ttg	ctg	2147
377	Lys	Tyr	Thr	Glu	Tyr	Glu	Ala	Arg	Thr	Lys	Ala	Ile	Glu	Thr	Leu	Leu	
378				240					245					250			
380	tac	gag	cga	ggg	ctc	atc	acg	ccc	gcc	gcg	gtc	gac	cga	gtc	gtt	tcg	2195

Use of n and/or Xaa has been detected in the Sequence Listing.
Review the Sequence Listing to insure a corresponding
explanation is presented in the <220> to <223> fields of
each sequence using n or Xaa.

VERIFICATION SUMMARY

PATENT APPLICATION: US/09/869,142

DATE: 07/09/2001

TIME: 10:10:44

Input Set : A:\PTO.txt

Output Set: N:\CRF3\07092001\I869142.raw

L:11 M:270 C: Current Application Number differs, Replaced Current Application No
L:11 M:271 C: Current Filing Date differs, Replaced Current Filing Date
L:607 M:341 W: (46) "n" or "Xaa" used, for SEQ ID#:6
L:617 M:341 W: (46) "n" or "Xaa" used, for SEQ ID#:6